Specificity of CBT for Depression: A Contribution from Multiple Treatments Meta-analyses(Dissertation_全文)

AUTHOR(S):

Honyashiki, Mina

Specificity of CBT for depression: a contribution from multiple treatments meta-analyses

Mina Honyashiki1

Toshi A. Furukawa1

Hisashi Noma2

Shiro Tanaka3

Peiyao Chen1

Kayoko Ichikawa4

Miki Ono5

Rachel Churchill6

Vivien Hunot6

Deborah M. Caldwell7

1Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan

2Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan

3Department of Pharmacoepidemiology, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan

4Department of Health Information, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan

5Department of Psychiatry, Kyoto University Graduate School of Medicine, Kyoto, Japan

6Academic Unit of Psychiatry, School of Social and Community Medicine, University of Bristol, Bristol, UK

7School of Social and Community Medicine, University of Bristol, Bristol, UK

ABSTRACT

The ñDodo bird verdict,ò which claims that all psychotherapies are equally effective, has been a source of bewilderment and intense controversy among psychiatrists and psychologists. To examine this issue, we focused on cognitive-behavior therapy (CBT) and applied the newly developed review method known as multiple treatments meta-analysis (MTM). We identified randomized controlled trials comparing CBT against a psychological placebo (PP) and/or no treatment (NT) controls during the acute phase treatment of adults with depression. A random-effects MTM was conducted within a Bayesian framework. All the analyses were performed on an intention-to-treat basis. The MTM of the evidence network from 18 studies (39 treatment arms, 1153 participants) revealed that CBT was significantly more likely to yield a response than NT (OR = 2.24, 1.32 to 3.88) and that CBT was nominally, but not significantly, superior to PP (OR = 1.30, 0.53 to 2.94), which in turn was superior to NT (OR = 1.73, 0.67 to 4.84). The intervention effects in MTM were associated with the number of sessions, and the specificity of CBT increased as the number of sessions increased. The specific component of CBT was estimated to constitute 50.4% (19.7 to 85.0) when CBT was given for 10 or more sessions. Despite the quantitatively and qualitatively limited body of randomized evidence examining this issue, the present study strongly suggested a non-null specific component of CBT when given for an adequate length.

KEYWORDS

Multiple treatments meta-analysis

Cognitive behavior therapy

Dodo bird verdict

Common factor

Specific factor

INTRODUCTION

It was Rosenzweig (1936) who first conceptualized psychotherapy as consisting of (i) common (non-specific) factors found in many different treatment approaches, and (ii) specific factors proper to a particular treatment method and theory. This conceptualization later paved the way for Rosenthal and Frankôs proposal of placebo psychotherapy, modeling pill placebo control in drug therapy trials, to establish the specific effectiveness of psychotherapies (Rosenthal & Frank, 1956). They wrote in 1956: ñé improvement under a special form of psychotherapy cannot be taken as evidence for (a) correctness of the theory on which it is based or (b) efficacy of the specific technique used, unless improvement can be shown to be greater than or qualitatively different from that produced by [é] a nonspecific form of psychotherapy.ò

The ensuing research efforts, however, have largely resulted in disappointing findings that are known as the Dodo bird verdict, which essentially states that all psychotherapies are equally effective (Baardseth et al., 2013; Luborsky et al., 2002; Luborsky & Singer, 1975; Smith & Glass, 1977; Wampold et al., 1997). The term originated from Rosenzweigôs citation from Lewis Carroll's novel ñAlice's Adventures in Wonderland,ò in which the characters get wet and have to dry themselves and the Dodo bird calls for a competition to run around the lake. When asked who won, the Dodo bird declares, ñEverybody has won, and all must have prizesò (Rosenzweig, 1936). The effectiveness of psychotherapies are thus postulated to be due to common factors, which include expectancy, relationship (empathy, warmth, alliance), and an explanatory framework (Greenberg & Newman, 1996; Omer & London, 1989).

However, the seminal papers cited above are subject to one or more of the following conceptual and methodological weaknesses.

1. As rightly criticized by Chambless and her colleagues (Chambless, 2002; Siev, Huppert, & Chambless, 2010), the authors of these papers (Baardseth et al., 2013; Luborsky et al.,

2002; Luborsky & Singer, 1975; Smith & Glass, 1977; Wampold et al., 1997) amalgamated very different comparisons for extremely diverse conditions among a wide spectrum of participants ranging from worried normal to psychotic inpatients. Their pooled effect size is therefore clinically uninterpretable. No one would choose his/her cancer therapy based on a meta-analysis of all therapies including all drugs, surgeries and radiation therapies for all stages of cancers of any histopathology and in any organ in the body.

2. Their dismissal of the obtained pooled effect size of 0.20 as small and clinically insignificant is factually and theoretically mistaken. First, one third of established and acknowledged interventions in both medicine and psychiatry have effect sizes smaller than 0.3 in comparison with a placebo (Leucht, Hierl, Kissling, Dold, & Davis, 2012). How can one expect a larger effect size when comparing active treatments? Second, an effect size of 0.20 corresponds with a number needed to treat (NNT) of around 15 for control event rates between 20%-50% (T. Furukawa, 1999). A common mental disorder often has a 12-month prevalence of 1% to 5%, which would translate into two to ten million sufferers per year in the USA alone; a therapy with an NNT of 15 could thus bring about 200,000 to 1,000,000 additional responses or remissions per year that an alternative therapy cannot achieve. This is not meaningless by any humane measure.
3. They base their arguments on the point estimate and ignore the uncertainties around it. In fact, the 95% confidence interval of their obtained effect size is very wide, surpassing 0.50, which signifies a moderate effect according to Cohenôs rule of thumb (Cohen, 1988) and may, in fact, be more powerful than more than half of the established and currently practiced medical interventions (Leucht et al., 2012). The correct statistical interpretation of the obtained pooled effect size in these studies should be: no firm evidence to exclude

neither clinically powerful difference in effect or no difference in effect, and not evidence of no clinically meaningful difference in effect.

4. It is most surprising that these meta-analyses are not based on a systematic search of all available evidence on a particular clinical topic, in view of the disconcerting magnitude of publication bias that has become widely known (Dickersin, 1990; Song, Eastwood, Gilbody, Duley, & Sutton, 2000). For example, Wampold and colleaguesô reviews limited their search to four English journals only (Ahn & Wampold, 2001; Wampold et al., 1997). Luborsky based their analyses on, alas, ñour collection of meta-analysesò (Luborsky et al., 2002).

On the other hand, there have also been attempts to refute the Dodo bird verdict by quantifying the specific vs. non-specific components in the effectiveness of psychotherapies, the most well-known of which is the one by Lambert and Barley (2001). Based on ña subset of more than 100 studies that provided statistical analyses of the predictors of outcomeò they concluded that specific techniques explained 15% of the total improvement in psychotherapy, the remaining being explained by common factors (30%), expectancy (15%) and extra therapeutic change (40%). Stevens, Hynan, and Allen (2000) were more specific: they calculated effect sizes for 80 outcome studies that each contained no treatment, a common factor, and treatment groups. The effect size in terms of symptom improvement was 0.58 for treatment vs. no treatment, which then was roughly additive of that between treatment and the common factor (0.26) and that between the common factor and no treatment (0.35). Bowers and Clum (1988) did a similar analysis for behavior therapy by performing a meta-analysis of studies that had both a placebo condition and a no treatment condition: the overall effect size of the treatment was 0.76, of which 0.55 was specific and 0.21 was non-specific. Barker, Funk, and Houston (1988) limited

themselves to credible placebo controls and found that the overall effect size of the treatment was 1.06, of which 0.55 was specific and 0.47 was non-specific. In other words, of the effectiveness of psychotherapies over no treatment, the percentage contributed by specific factors ranged widely, with values of 25%, 45%, 72%, and 52%, respectively. None of these figures may be clinically meaningless, but unfortunately all these reviews are subject to some or all of the criticisms described above.

Therefore, it is timely to ask how much specific vs. non-specific components there are in the effectiveness of a specific psychotherapy for a well-delineated clinical condition using a modern systematic review methodology. The current study represents a secondary analysis of the Cochrane systematic reviews of six major psychotherapy schools for depression in adults (Hunot et al., 2013; Shinohara et al., 2013). The six schools included behavior therapies, cognitive-behavior therapies, third-wave cognitive therapies, psychodynamic therapies, humanistic therapies and integrative therapies. In order to quantitatively assess the specific vs. non-specific components, the present study focuses on a triangular comparison between cognitive-behavior therapies (CBT), which were the most thoroughly researched of the six schools, and a psychological placebo (PP) and no treatment (NT). We also applied a new meta-analysis technique, known as multiple treatments meta-analysis (MTM) or network meta-analysis (Higgins & Whitehead, 1996), to this triangular comparison to combine the direct and indirect comparisons contained therein, so that we can make the maximal use of the available randomized evidence.

METHODS

Criteria for considering studies for this review

We included only randomized controlled studies comparing CBT with PP and/or NT in the acute phase treatment of adults with depression. Quasi-randomized studies, such as those using allocation by day of the week, date of birth, or alternate allocations, were not eligible because a lack of allocation concealment leads to overestimation (Schulz, Chalmers, Hayes, & Altman, 1995). Both open and single-blinded (assessor-blinded) studies were eligible, as it is impossible to blind the therapists or participants in psychotherapy trials.

Depression could either be defined as unipolar major depression according to any of the operationalized diagnostic criteria (Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, ICD-10) or as scoring above the accepted threshold of a validated depression screening instrument. Studies focusing on chronic or treatment-resistant depression were excluded.

CBT includes cognitive therapy (Beck, Rush, Shaw, & Emery, 1979), rational emotive behavior therapy (Ellis, 1979), problem-solving therapy (D'Zurilla & Goldfried, 1971), self-control therapy (Fuchs & Rehm, 1977), Coping with Depression Course (Lewinsohn, Antonuccio, Breckenridge, & Teri, 1984) and others that use both cognitive and behavioral skills for the treatment of depression.

PP is defined as an experimental condition used in an attempt to control for non-specific factors. The criteria for a control condition to be regarded PP were as follows: (1) intervention is regarded as lacking active components by researchers in a trial but is explained as active to the participants; (2) the number and duration of the face-to-face session is equivalent with active treatment in the same study and; (3) the qualification of the therapists is equivalent to that for the active treatment. We did not include pill placebo controls because they control for the regression towards the mean, the natural course and treatment expectancy but not the common

therapeutic factors of psychotherapy (Hollon & DeRubeis, 1981).

NT consists of patients who did not receive either active or non-specific interventions. This control condition controls for the regression towards the mean and the natural course of the condition. We did not include waiting list controls, which are often used in psychotherapy research, among the NT controls.

Study selection and data extraction

To identify relevant studies, we searched two clinical trial registries created and maintained by the Cochrane Depression, Anxiety and Neurosis Group (CCDAN), the CCDANCTR-Studies and CCDANCTR-References, supplemented by corresponding searches in CINAHL, PSYINDEX, and reference searches. The details of the search strategies for these registries can be found on the Cochrane Collaboration Depression, Anxiety and Neurosis Groupôs webpage (http://ccdan.cochrane.org/). The most recent updated search for this review was done in February 2012. The quality ratings were operationalized, and studies were categorized into either a low risk of bias, a high risk of bias, or an unclear risk of bias for each domain. All the assessments were performed by two independent review authors, and disagreements were resolved by discussion between two authors and, where necessary, in consultation with a third author. Missing information was sought by contacting the original authors, whenever possible.

Outcome measures

Acute treatment was defined as an 8-week treatment in the analyses. If 8-week data were not available, we used data ranging between 4 to 16 weeks, and the time point given in the original study as the study endpoint was given preference.

Response was our pre-defined primary outcome, as this allows the inclusion of all dropouts and

thus enables a conservative estimate of the treatment effect according to the intention-to-treat principle. We defined response as the proportion of patients who showed a reduction of at least 50% from the baseline score on the Hamilton Rating Scale for Depression (HAM-D), the Montgomery-Asberg Depression Rating Scale (MADRS), or any other validated depression scale at the above-defined time point. If the original authors reported several outcomes, we gave preference to the BDI for a self-rating scale and the HAM-D for an observer-rating scale. Observer-rated scales were preferred to self-reported scales.

Intention-to-treat analyses were based on the total number of randomly assigned participants, irrespective of how the original study investigators analyzed the data, by assuming that all dropouts were non-responders. For studies in which the exact numbers of participants who had responded were not reported, but the means and standard deviations for continuous depression scales were reported, the number of responders was calculated using a validated imputation method (da Costa et al., 2012; T. A. Furukawa, Cipriani, Barbui, Brambilla, & Watanabe, 2005).

Analysis

Multiple treatments meta-analyses, and examination of inconsistency/heterogeneity

We conducted multiple treatments meta-analyses. To ensure that the network was connected, a network diagram was constructed. Random-effects MTM, allowing for the heterogeneity of treatment effects across studies, was conducted in a Bayesian framework using OpenBUGS 3.2.1. These methods combine direct and indirect evidence for all three pairs of treatments. A key assumption of MTM is that of consistency, i.e., that direct and indirect evidence do not disagree beyond chance. In the first instance, one should ensure that the subsets of trials forming the network are similar in factors which could modify the treatment effect. Where feasible, consistency should also be statistically evaluated. Here, we used the posterior mean of the

residual deviance as a global goodness of fit statistic to assess consistency. In a well-fitting model. the residual deviance should be close to the number of data points. In case with considerable inconsistency, we investigated the possible sources.

Quantifying specific vs. non-specific components

The relative contributions of specific effects and non-specific effects were estimated by dividing log ($OR_{CBT,PP}$) or log ($OR_{PP,NT}$) by log ($OR_{CBT,NT}$), where $OR_{X,Y}$ represents the odds ratio of treatment X over treatment Y.

Publication bias and sensitivity analyses

To assess publication bias, we drew funnel plots for pair-wise comparisons if the number of studies contributing to that comparison was 10 or greater. To examine if the obtained results were preserved when we limited the included studies to only high-quality ones, we had planned a priori to examine the following variables: risk of biases (limiting to trials with a low risk of bias at allocation concealment, blinding of assessor, and treatment fidelity), included disorders, and response imputation.

Meta-regression

The following sources of possible clinical heterogeneity, which had been listed a priori, were examined as effect modifiers in network meta-analyses: number of sessions, group vs. individual format, baseline depression severity, and concomitant pharmacotherapy.

RESULTS

Selection and inclusion of studies

Out of 6710 studies identified through an electronic search and reference search, 195 full-text

articles were retrieved, of which 18 studies (comprising 39 treatment arms, and 1153 participants) satisfied the eligibility criteria for the present study (Figure 1).

Characteristics of the included studies

Figure 2 shows the network of evidence comparing CBT, PP, and NT. The characteristics of the included studies are listed in Table 1. The contents of the PP conditions are listed in Table 2. Two of the 18 studies had two CBT arms. Five of the 18 studies used an individual format for CBT or PP, 11 studies used a group format, and the remaining two used both formats. The number of sessions ranged from 4 to 12 sessions. Ten of the 15 studies allowed concomitant pharmacotherapy, while five studies did not. Only two studies used an observer scale (HAMD) as an outcome measure, while the other 16 studies used a self-rating scale (BDI). The mean baseline severity on the BDI was minimal (14-19) in one study, mild (20-28) in 14 studies, and moderate (>28) in one study. The quality of the included studies varied but was generally moderate. Ten studies reported adequate allocation concealment. One out of two studies using an objective scale reported the blinding of the assessors. Three studies reported fidelity monitoring for CBT or PP. Twelve studies included patients with major depressive disorder diagnosed according to operationalized diagnostic criteria, while the remaining six included patients scoring above the accepted threshold of a validated depression screening instrument. We had to use the imputed response rates based on the continuous severity score at the end of treatment in 16 studies. All but one study provided data on the numbers of randomized patients. We used the number of participants assessed at the end of treatment as the denominator for the remaining study.

Pair-wise meta-analyses

We conducted CBT vs. PP and CBT vs. NT pair-wise meta-analyses (Table 3). These analyses showed that CBT was significantly more effective than NT in bringing about a response. The CBT vs. PP comparison was not significant. Overall, the heterogeneity was moderate, although for all comparisons the 95% CI included values that showed very high or no heterogeneity, reflecting the small number of included studies for each pair-wise comparison.

Multiple treatment meta-analyses and examination of inconsistency/heterogeneity

The consistency model provided an adequate fit to the data, with a posterior mean residual deviance of 37.8 for 37 data points, although an index of heterogeneity (the median between-trials standard deviation) was relatively high (ů = 0.70). Table 4 summarizes the results of the MTM. CBT was significantly superior to NT. CBT was not significantly different from PP, nor was PP from NT.

Publication bias and sensitivity analyses

We drew a funnel plot for the primary outcome of the studies comparing CBT and NT. Eggerôs test was not significant ($P = 0.34$). For other comparisons, the number of comparisons was too small for a funnel plot.

There were not enough studies to conduct MTM for sensitivity analyses, so we only conducted pair-wise meta-analyses. Among them, limiting the studies to high-quality trials did not change the overall results (see Table 3).

Meta-regression

We conducted meta-regressions for MTM to examine the effects of selected covariates on

efficacy. The association between the treatment effect and the number of sessions was significant (slope: -0.21; 95%CrI: -0.42 to -0.002). We found no indication that the treatment efficacy was significantly associated with the baseline depression severity according to the BDI (slope: -0.05; 95%CrI: -0.21 to 0.10), nor did we find an association between the effect size and the CBT format (slope: -0.04; 95%CrI: -1.28 to 1.18) or concomitant pharmacotherapy (slope: -0.52; 95%CrI: -1.56 to 0.45).

Figure 3 shows the estimated relationship between the number of sessions and the specificity of CBT. Table 4 presents a post-hoc meta-regression dichotomizing the number of session into ñÓ10ò and ñ<10ò. The specific component now contributed 50.4% (95%CrI: 19.7% to 85.0%) of the total efficacy of CBT over NT when the number of sessions was 10 or over. The interaction was qualitative (Table 4), suggesting that CBT is specifically beneficial only if it is given in 10 or more sessions.

DISCUSSION

A systematic comprehensive search of the literature yielded a network of evidence of 18 studies (comprising 39 arms, and 1153 patients) comparing CBT, PP, and NT. The MTM of the evidence network was consistent, revealing that CBT was significantly more likely to yield a response than NT (OR = 2.24, 1.32 to 3.88) and that CBT was nominally, but not significantly, superior to PP (OR = 1.30, 0.53 to 2.94), which in turn was superior to NT (OR = 1.73, 0.67 to 4.84). For all the comparisons, the credible intervals were relatively wide because of the lack of power. The specificity of CBT was estimated to constitute 35.0% (-99.5% to 180.3%) of its efficacy over NT.

Pooling all available evidence, the estimate for the specificity of CBT had an extremely wide credible interval. In other words, overall, the currently available best evidence was compatible with both the no specificity hypothesis, i.e., the Dodo bird verdict (Baardseth et al., 2013; Luborsky et al., 2002; Luborsky & Singer, 1975; Smith & Glass, 1977; Wampold et al., 1997), as well as all foregoing point estimates ranging between 25% through 72%(Barker et al., 1988; Bowers & Clum, 1988; Lambert & Barley, 2001; Stevens et al., 2000). However, post-hoc exploratory analyses revealed that CBT of adequate length had a specificity component of about 50%, with a 95% credible interval between 20% and 85%. We may now assume, with some confidence, that CBT has a non-zero specific component in the treatment of depression in adults.

There is now corollary evidence to suggest that the Dodo bird verdict is not universally operative. Critical incident stress debriefing is a form of crisis counseling aimed at preventing the development of posttraumatic stress disorder. It is typically delivered to a group of trauma survivors in a single 1ï 3-hour session that takes place within one week of the trauma event. Although it does contain many common factors, such as empathic listening by experts in the field with credible explanatory models, specific factors appear to be at work leading to null to harmful results (Rose, Bisson, Churchill, & Wessely, 2002; van Emmerik, Kamphuis, Hulsbosch, & Emmelkamp, 2002). Cottraux et al. demonstrated that cognitive therapy and exposure therapy may have differential degrees of effectiveness on obsessive-compulsive disorder (OCD), with the former having greater effects on depression and anxiety and the latter having greater effects on intrusive thoughts and OCD symptoms. They also reported some analyses showing that the amount of specific effects increases from post-treatment to follow-up, which could indicate that the post-treatment results are more strongly influenced by common factors, while follow-up

assessments can reflect more specific components (Cottraux et al., 2001).

The number of included studies may appear limited in comparison with some recent systematic reviews of CBT for depression (Barth et al., 2013; Jakobsen, Hansen, Storebo, Simonsen, & Gluud, 2011), but our objective was not to perform a systematic review of CBT in general but to ask a focused question regarding the specificity of CBT by performing a network meta-analysis, for which the homogeneity and consistency of the included interventions and populations were more important than for traditional pairwise meta-analyses. We therefore focused on face-to-face CBT, with patients who were diagnosed as having acute depression according to operationalized diagnostic criteria or by scoring above the accepted threshold of a validated depression screening instrument. We also did not include behavior therapy or third-wave CBT in order to focus on narrowly defined CBT. We excluded studies if they employed protocolized pharmacotherapy in conjunction with CBT. Neither did we include the waiting list control, often used in psychotherapy research, as an NT control because there is a growing suspicion that the waiting list control may be differentiated from the no treatment condition (Watanabe, Hunot, Omori, Churchill, & Furukawa, 2007). We further limited PP to interventions that were regarded as lacking an active component by researchers in the trial but that were explained as having an active component to the participants. We did not consider so-called counseling or supportive psychotherapy as PP because we believe these techniques have active components and should be classified as an active treatment. We adopted this narrow definition of PP in order to avoid bias due to researcher allegiance. All in all, out of the 128 studies found in the original study selection, we were only able to include 18 studies comparing CBT with PP and/or NT during the acute phase treatment of adults with depression (Figure 1).

Several caveats are in order before we conclude. First, despite our systematic and comprehensive search of the literature, we were able to include only a relatively small number of studies. Thus, for example, although the network meta-regression revealed that the specific component of CBT may constitute half of its efficacy when CBT was given for 10 or more sessions, it ought to be noted that only 5 of the 18 studies had 10 or more sessions. Secondly, the evidence was not only quantitatively, but also qualitatively less than desirable. Allocation concealment was reported to be adequate in only three studies, and assessor blinding was reported in only one of the 18 studies. Furthermore, only three studies examined treatment fidelity in a satisfactory manner, and the response rates had to be imputed from the reported continuous outcomes in all but two studies. The results, however, were robust to sensitivity analyses. Thirdly, the heterogeneity of evidence network among CBT, PP, and NT, measured in terms of the median between-trial standard deviation, was relatively large when compared with the estimated effect sizes between the treatment arms. The heterogeneity coupled with the small sample size may have limited the power to detect relatively weak but important effect modifiers. We were not able to conduct many of the pre-planned sensitivity analyses, and where we were able to perform such analyses, they may have lacked an adequate power. However, when we included characteristics of the trials as effect modifiers and when the heterogeneity arising from the number of sessions was accounted for, the median between-trial standard deviations decreased. Last, but not least, our analytical model supposes a simple additive relationship between specific and non-specific components. However, it is imaginable that some interaction may exist between the two types of components: for example, if a treatment is very effective from its beginning, this would increase the patientsô expectations for a positive outcome and hence would increase the placebo effect, but this can occur only in the treatment group. We would need better-designed studies, possibly with multiple control conditions with differential

intensities, to detect such interactions.

On the other hand, the strengths of the present study may be as follows. First and foremost, we started with a well-formulated and well-focused clinical question to examine the specificity of a well-delineated intervention, i.e. CBT, for a specific clinical condition, i.e. acute phase treatment of depression in adults. Secondly, we followed the Cochrane review methodology. Comprehensive literature searches were conducted so as to minimize publication bias (Egger, Juni, Bartlett, Holenstein, & Sterne, 2003). Detailed manuals were prepared to guide the selection and data extraction of studies in duplicates. We also examined possible sources of bias and conducted analyses following an intention-to-treatment principle as closely as possible. Thirdly, the use of MTM has enabled us to examine the consistency of the totality of evidence surrounding CBT, PP, and NT and to derive the most precise estimate of the specific component of CBT possible based on randomized evidence, while adjusting for possible effect modifiers. Thus, the main weaknesses of previous reviews, namely the unfocused inclusion of participants and interventions, the lack of systematic searches, and the small effect sizes with wide 95% confidence intervals, have all been addressed in this study.

In conclusion, the present study represents the most up-to-date and comprehensive summary for the specificity hypothesis of CBT for depression. Despite the quantitatively and qualitatively limited body of randomized evidence examining this issue, the present study suggested a non-null specific component for one form of psychotherapy for one particular disorder. Future studies are needed to assess the specificity of CBT and other well-defined psychotherapies of adequate length and of satisfactory quality for various psychiatric disorders and psychological problems. Such psychotherapies, when they do exist, should be given preference in the

provision and training of psychotherapies. The Dodo bird verdict is on the verge of extinction.

DISCLOSURES

TAF has received honoraria for speaking at CME meetings sponsored by Asahi Kasei, Eli Lilly, GlaxoSmithKline, Mochida, MSD, Otsuka, Pfizer, Shionogi and Tanabe-Mitsubishi. He is diplomate of the Academy of Cognitive Therapy. He has received royalties from Igaku-Shoin, Seiwa-Shoten and Nihon Bunka Kagaku-sha. He is on advisory board for Sekisui Chemicals and Takeda Science Foundation. All the other authors have no conflicts of interest to declare. This study was supported in part by Grant-in-Aid by the Ministry of Health, Labour and Welfare to TAF. The study has not been presented at any meeting.

REFERENCES

Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. *Journal of Counseling Psychology, 48*, 251-257.

Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., et al. (2013). Cognitive-behavioral therapy versus other therapies: redux. *Clin Psychol Rev, 33*(3), 395-405.

Barker, S. L., Funk, S. C., & Houston, B. K. (1988). Psychological Treatment Versus Nonspecific Factors - a Meta-Analysis of Conditions That Engender Comparable Expectations for Improvement. *Clinical Psychology Review, 8*(6), 579-594.

Barth, J., Munder, T., Gerger, H., Nuesch, E., Trelle, S., Znoj, H., et al. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Med, 10*(5), e1001454.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive Therapy of Depression*. New York: Guilford Press.

Besyner, K., J. (1979). The comparative efficacy of cognitive and behavioral treatments of depression: A multi-assessment approach. *A Thesis Submitted to the Graduate Faculty of Texas Tech University*

Bowers, T. G., & Clum, G. A. (1988). Relative Contribution of Specific and Nonspecific Treatment Effects - Meta-Analysis of Placebo-Controlled Behavior-Therapy Research. *Psychological Bulletin, 103*(3), 315-323.

Chambless, D. (2002). Beware the dodo bird: The dangers of overgeneralization. *Clinical Psychology: Science and Practice, 9*, 13-16.

Cohen, J. (1988). *Statistical Power Analysis in the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.

Cottraux, J., Note, I., Yao, S. N., Lafont, S., Note, B., Mollard, E., et al. (2001). A randomized controlled trial of cognitive therapy versus intensive behavior therapy in obsessive compulsive disorder. *Psychother Psychosom, 70*(6), 288-297.

D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of abnormal psychology, 78*(1), 107-126.

da Costa, B. R., Rutjes, A. W., Johnston, B. C., Reichenbach, S., Nüesch, E., Tonia, T., et al. (2012). Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. *International Journal of Epidemiology, 41*(5), 1445-1459.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama, 263*(10), 1385-1389.

Dowrick, C., Dunn, G., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., Lehtinen, V., et al. (2000). Problem solving treatment and group psychoeducation for depression: multicentre

randomised controlled trial. Outcomes of Depression International Network (ODIN) Group. *BMJ, 321*(7274), 1450-1454.

Dunn, G., Maracy, M., Dowrick, C., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., et al. (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. *Br J Psychiatry, 183*, 323-331.

Egger, M., Juni, P., Bartlett, C., Holenstein, F., & Sterne, J. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess, 7*(1), 1-76.

Ellis, A. (1979). *Theoretical and Empirical Foundations of Rational-Emotive Therapy*. Monterey: Brooks/Cole.

Faramarzi, M., Alipor, A., Esmaelzadeh, S., Kheirkhah, F., Poladi, K., & Pash, H. (2008). Treatment of depression and anxiety in infertile women: cognitive behavioral therapy versus fluoxetine. *J Affect Disord, 108*(1-2), 159-164.

Fuchs, C. Z., & Rehm, L. P. (1977). A self-control behavior therapy program for depression. *Journal of consulting and clinical psychology, 45*(2), 206-215.

Furukawa, T. (1999). From effect size into number needed to treat. *Lancet, 353*(9165), 1680-1680.

Furukawa, T. A., Cipriani, A., Barbui, C., Brambilla, P., & Watanabe, N. (2005). Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol, 20*(1), 49-52.

Greenberg, L. S., & Newman, F. L. (1996). An approach to psychotherapy change process research: Introduction to the special section. *Journal of Consulting and Clinical Psychology, 64*(3), 435-438.

Hamamci, Z. (2006). Integrating psychodrama and cognitive behavioral therapy to treat

moderate depression. *The Arts in Psychotherapy, 33*(3), 199-207.

Hamdan-Mansour, A. M., Puskar, K., & Bandak, A. G. (2009). Effectiveness of cognitive-behavioral therapy on depressive symptomatology, stress and coping strategies among Jordanian university students. *Issues Ment Health Nurs, 30*(3), 188-196.

Hegerl, U., Hautzinger, M., Mergl, R., Kohnen, R., Schutze, M., Scheunemann, W., et al. (2010). Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. *Int J Neuropsychopharmacol, 13*(1), 31-44.

Higgins, J. P., & Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Stat Med, 15*(24), 2733-2749.

Hollon, S. D., & DeRubeis, R. J. (1981). Placebo-psychotherapy combinations: inappropriate representations of psychotherapy in drug-psychotherapy comparative trials. *Psychol Bull, 90*(3), 467-477.

Hunot, V., Moore Theresa, H. M., Caldwell Deborah, M., Furukawa Toshi, A., Davies, P., Jones, H., et al. (2013). 'Third wave' cognitive and behavioural therapies versus other psychological therapies for depression. *Cochrane Database of Systematic Reviews,* (10). Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008704.pub2/abstract. doi:10.1002/14651858.CD008704.pub2

Jakobsen, J. C., Hansen, J. L., Storebo, O. J., Simonsen, E., & Gluud, C. (2011). The effects of cognitive therapy versus 'no intervention' for major depressive disorder. *PLoS One, 6*(12), e28299.

Kelly, M. L. (1982). Rational emotive therapy versus Lewinsohnian based approaches to the

treatment of depression. *Dissertation Abstracts International, 43*(6-B).

Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy, 38*(4), 357-361.

Leucht, S., Hierl, S., Kissling, W., Dold, M., & Davis, J. M. (2012). Putting the efficacy of psychiatric and general medicine medication in perspective: a review of meta-analyses. *British Journal of Psychiatry, 200*, 97-106.

Lewinsohn, P. M., Antonuccio, D. O., Breckenridge, J. S., & Teri, L. (1984). *The Coping With Depression Course: A psychoeducational intervention for unipolar depression*. Eugene, OR: Castalia.

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., et al. (2002). The dodo bird verdict is alive and well --- mostly. *Clinical Psychology: Science and Practice, 9*, 2-12.

Luborsky, L., & Singer, B. (1975). Comparative studies of psychotherapies. Is it true that "everywon has one and all must have prizes"? *Archives of general psychiatry, 32*(8), 995-1008.

Miranda, J., Chung, J. Y., Green, B. L., Krupnick, J., Siddique, J., Revicki, D. A., et al. (2003). Treating depression in predominantly low-income young minority women: a randomized controlled trial. *JAMA, 290*(1), 57-65.

Omer, H., & London, P. (1989). Signal and noise in psychotherapy. The role and control of non-specific factors. *Br J Psychiatry, 155*, 239-245.

Propst, L. R. (1980). The comparative efficacy of religious and nonreligious imagery for the treatment of mild depression in religious individuals. *Cognitive Therapy and Research, 4*(2), 167-178.

Rose, S., Bisson, J., Churchill, R., & Wessely, S. (2002). Psychological debriefing for

preventing post traumatic stress disorder (PTSD). *Cochrane database of systematic reviews*(2), CD000560.

Rosenthal, R., & Frank, J. D. (1956). Psychotherapy and the placebo effect. *Psychological Bulletin, 53*(4), 294-302.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry, 6*, 412-415.

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Jama, 273*(5), 408-412.

Serfaty, M. A., Haworth, D., Blanchard, M., Buszewicz, M., Murad, S., & King, M. (2009). Clinical effectiveness of individual cognitive behavioral therapy for depressed older people in primary care: a randomized controlled trial. *Arch Gen Psychiatry, 66*(12), 1332-1340.

Shinohara, K., Honyashiki, M., Imai, H., Hunot, V., Caldwell Deborah, M., Davies, P., et al. (2013). Behavioural therapies versus other psychological therapies for depression. *Cochrane Database of Systematic Reviews,* (10). Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008696.pub2/abstract. doi:10.1002/14651858.CD008696.pub2

Siev, J., Huppert, J., & Chambless, D. (2010). Treatment specificity for panic disorder: A reply to Wampold, Imel, and Miller (2009). *Behavior Therapist, 33*, 12-14.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *The American psychologist, 32*(9), 752-760.

Song, F., Eastwood, A. J., Gilbody, S., Duley, L., & Sutton, A. J. (2000). Publication and related biases. *Health Technol Assess, 4*(10), 1-115.

Stevens, S. E., Hynan, M. T., & Allen, M. (2000). A meta-analysis of common factor and specific treatment effects across the outcome domains of the phase model of psychotherapy. *Clinical Psychology-Science and Practice, 7*(3), 273-290.

van Emmerik, A. A., Kamphuis, J. H., Hulsbosch, A. M., & Emmelkamp, P. M. (2002). Single session debriefing after psychological trauma: a meta-analysis. *Lancet, 360*(9335), 766-771.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "All must have prizes.". *Psychological Bulletin, 122*, 203-215.

Watanabe, N., Hunot, V., Omori, I. M., Churchill, R., & Furukawa, T. A. (2007). Psychotherapy for depression among children and adolescents: a systematic review. *Acta Psychiatr Scand, 116*(2), 84-95.

Table 1. Selected characteristics of the included studies

Study	N of arms in:			N	Included disorders	Baseline BDI	Format	N of sessions	Con-comitant pharmaco-therapy	Outcome scale	Risk of Bias			Response imputed
	CBT	PP	NT								Allocation concealment	Blinding of assessors	Treatment fidelity	
Besyner1979 (Besyner, 1979)	1	1		20	Other	24.9	Grp	4	Unclear	BDI	Unclear	High	Unclear	Imputed
[a]Dowrick_Finland Rural1996 (Dowrick et al., 2000; Dunn et al., 2003)	1		1	50	MDD+	21.1	Ind	6	Allowed	BDI	Low	High	Unclear	Imputed
[a]Dowrick_Finland Urban1996 (Dowrick et al., 2000; Dunn et al., 2003)	1		1	47	MDD+	21.3	Ind	6	Allowed	BDI	Low	High	Unclear	Imputed
[a]Dowrick_Ireland UrbanRural1996 (Dowrick et al., 2000; Dunn et al., 2003)	1		1	38	MDD+	23	Grp	8	Allowed	BDI	Low	High	Unclear	Imputed
[a]Dowrick_Norway Rural1996 (Dowrick et al., 2000; Dunn et al., 2003)	1		1	61	MDD+	19.2	Grp	8	Allowed	BDI	Low	High	Unclear	Imputed
[a]Dowrick_Norway Urban1996 (Dowrick et al., 2000; Dunn et al., 2003)	1		1	67	MDD+	21	Grp	8	Allowed	BDI	Low	High	Unclear	Imputed
[a]Dowrick_Spain	1		1	30	MDD+	22	Ind	6	Allowed	BDI	Low	High	Unclear	Imputed

Urban1996 (Dowrick et al., 2000; Dunn et al., 2003)														
[a]Dowrick_UK Rural1996 (Dowrick et al., 2000; Dunn et al., 2003)	1		1	49	MDD+	26	Ind	6	Allowed	BDI	Low	High	Unclear	Imputed
[a]Dowrick_UK Urban1996 (Dowrick et al., 2000; Dunn et al., 2003)	2		1	84	MDD+	24.8	Ind/Grp	6/8	Allowed	BDI	Low	High	Unclear	Imputed
Faramarzi2008 (Faramarzi et al., 2008)	1		1	82	Other	19.9	Grp	10	No	BDI	Unclear	High	Unclear	Imputed
Fuchs1977 (Fuchs & Rehm, 1977)	1	1		[b]18	Other	NA	Grp	6	Unclear	BDI	Unclear	High	Unclear	Imputed
Hamamci2006 (Hamamci, 2006)	1		1	24	Other	28.4	Grp	11	No	BDI	Unclear	High	Unclear	Imputed
Hamdan-Mansour2009 (Hamdan-Mansour, Puskar, & Bandak, 2009)	1		1	84	Other	24.1	Grp	10	Unclear	BDI	Low	High	Low	Imputed
Hegerl2010 (Hegerl et al., 2010)	1	1		120	MDD+	NA	Grp	10	No	HAMD	Unclear	Unclear	Low	No
Kelly1982 (Kelly, 1982)	1	1		16	MDD+	25.4	Grp	6	Allowed	BDI	Unclear	High	Unclear	Imputed
Miranda2003 (Miranda et al., 2003)	1		1	179	MDD+	NA	Ind/Grp	8	No	HAMD	Low	Low	Unclear	No
Propst1980 (Propst, 1980)	2	1	1	47	Other	15.4	Grp	8	No	BDI	Unclear	High	Unclear	Imputed
Serfaty2009 (Serfaty et al., 2009)	1	1		137	MDD+	26.8	Ind	12	Allowed	BDI	Low	High	Low	Imputed

Abbreviations CBT: cognitive behavior therapies; PP: psychological placebo; NT: no treatment; BDI: Beck Depression Inventory; MDD+: Major depressive disorder diagnosed by operationalised diagnostic criteria

[a]Dowrick et al. (2000) reports nine independently conducted, albeit according to concerted protocols, RCTs. Two of these RCTs conducted in Ireland were reported in an amalgamated form in the definitive report (Dunn et al., 2003) and is therefore treated as one trial in this meta-analysis.

[b]For Fuchs and Rehm (1977), randomized N was not available. Instead we used number of participants assessed at the end of intervention.

Table 2. Description of psychological placebo conditions in each study

Study	Description of PP
Besyner1979	Nonspecific group: ñTherapist behavior was limited to reflection and clarification of verbal material and questioning to facilitate discussion. It may be argued that such procedures are akin to, if not identical with, those employed by Rogerian therapists. While the validity of this argument cannot be denied, it is the belief of this researcher that such procedures are considered to be minimally therapeutic.ò (page 70, line 10)
Fuchs1977	Nonspecific therapy: ñSession 1 began in the same way as the self-control procedure with introductions, collection of deposits, a review of confidentiality issues, and a 10-minute group interaction assessment procedure. As in the other groups, participants were given an information sheet and a general introduction to group therapy concepts, generally from a nondirective framework. From that point on and throughout the ensuing sessions, therapists in this condition attempted to elicit discussion of past and current problems, to encourage group interaction, and to reflect and clarify feelings in an empathic manner. Although therapists at times suggested simple exercises within the group to facilitate open discussion, they were specifically instructed neither to recommend out-of-therapy activity nor explicitly to teach behavioral principles. These sessions lasted approximately 2 hours weekly, as did self-control therapy sessions.ò (page 209, left column, line 24)
Hegerl2010	Guided self help group (GSG): ñIn the GSG , a supportive atmosphere was created, allowing the participants to communicate about their situation and daily life, but no psychotherapeutic intervention was allowed by the group leader.ò (page 33, right column, line 1)
Kelly1982	Nondirective group: òThe nondirective group served as a control group and met for the same amount of time as the other groups, but did not

	undergo their treatment procedures. Outside of behavior change strategies and cognitive strategies, the group was free to discuss any topics (e.g., support, jobs, etc.). All sessions, with the exception of the first, consisted of a review of the previous meetingôs topic and a discussion of issues the group members felt were important. The therapist behavior during all sessions was as consistent as possible. An attempt was made to provide all group members with maximum empathy and warmth.ò (page 41, line 10)
Propst1980	Therapist Contact plus Self-Monitoring: ñParticipants in this condition simply met for a discussion group and kept track of their daily mood. For homework they were to record items for group discussion on their mood cards. The content of the discussion was up to the participants, as the therapists participated as little as possible.ò (page 172, line 5)
Serfaty2009	Talking Control: ñClearly defined criteria for the TC group were used to prevent CBT from being delivered. Talking control therapy was developed during our feasibility work, and details are available from the authors. The therapists practiced delivering the TC in role plays with the supervisor so that difficult questions could be addressed. Dysfunctional beliefs were not challenged; however, the therapists were asked to show interest and warmth, encouraging participants to discuss neutral topics such as hobbies, sports, and current affairs. No advice or problem solving was given, and there was little focus on emotional issues. No suggestions for behavioral tasks were offered. So for example, if the patient said, ñMy daughter does not like me as she never comes to visit me,ò the therapist would ask, ñHow many children do you have?ò (page 1334, right column, line 8)

Table 3. Pair-wise meta-analyses and sensitivity analyses

	Pair wise meta-analyses		Allocation concealment		Blinding of assessors		Treatment fidelity		Included disorders		Response imputed	
	OR (95%CI)	n	OR (95%CI)	n	OR (95%CI)	n	OR (95%CI)	n	OR (95%CI)	n	OR (95%CI)	n
CBT vs NT	2.07 (1.35 to 3.18)	13	1.79 (1.18 to 2.71)	10	1.31 (0.67 to 2.52)	1	7.00 (2.31 to 21.19)	1	1.49 (1.03 to 2.15)	9	1.31 (0.67 to 2.52)	1
CBT vs PP	1.74 (0.79 to 3.83)	6	1.55 (0.84 to 2.83)	1	NA	0	2.54 (1.34 to 4.82)	2	2.11 (1.16 to 3.83)	3	4.89 (1.53 to 15.66)	1
PP vs NT	2.04 (0.40 to 10.55)	1	NA	0	NA	0	NA	0	NA	0	NA	0

n:number of included studies

Table 4. Odds ratios of response and specificity of CBT estimated in MTM and its meta-regression

		Meta-regression MTM	
	Overall MTM	< 10sessions	Ó10sessions
CBT vs NT	2.24	1.53	7.37
	(1.32 to 3.88)	(1.02 to 2.28)	(3.74 to 15.15)
CBT vs PP	1.30	0.55	2.71
	(0.53 to 2.94)	(0.27 to 1.20)	(1.42 to 5.33)
PP vs NT	1.73	2.72	2.72
	(0.67 to 4.84)	(1.28 to 5.76)	(1.28 to 5.76)
CBT specific component	35.0%	-159.6%	50.4%
	(-99.5 % to 180.3%)	(-958.4% to 90.6%)	(19.7% to 85.0%)

Numbers in parentheses represent 95% credible intervals.

Figure 1. Flowchart for selection of studies

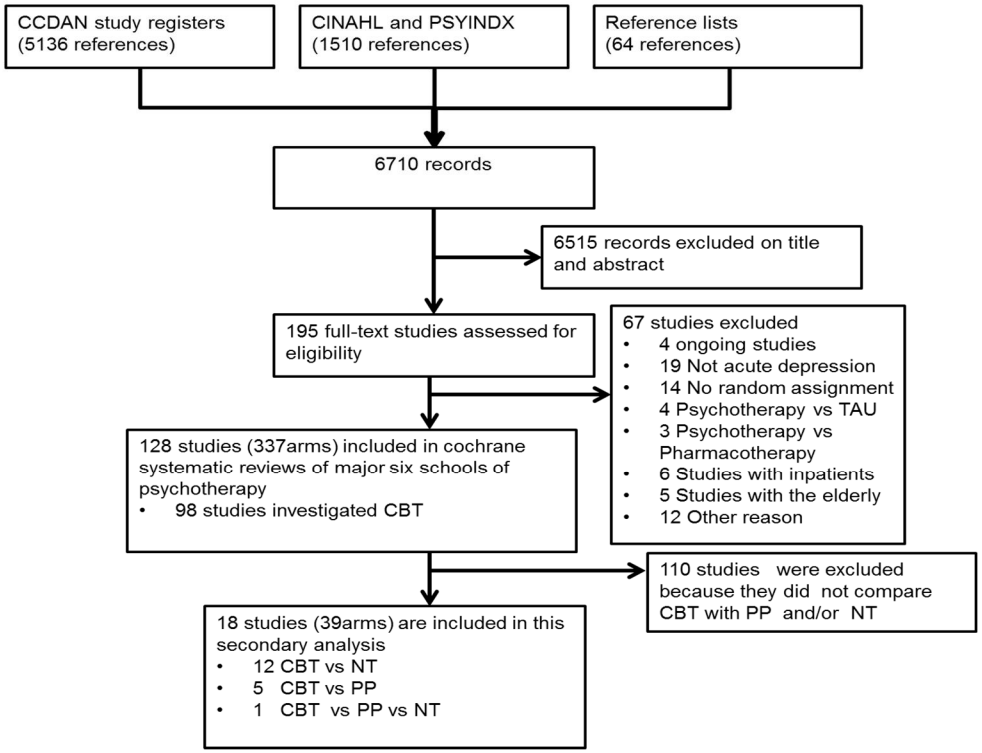
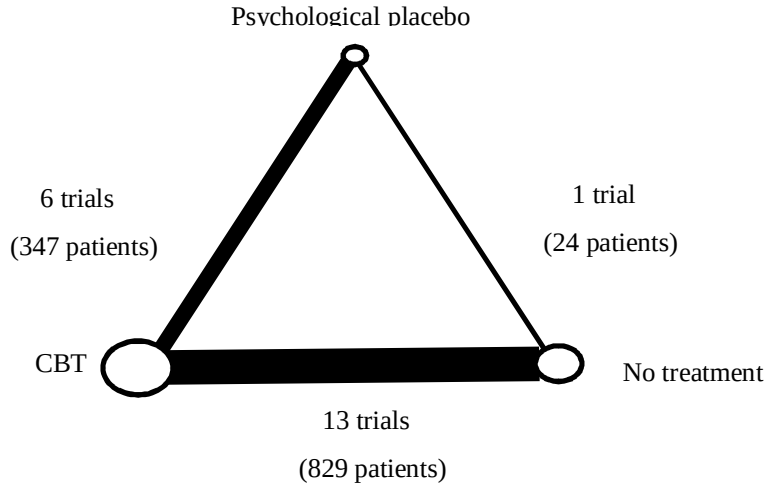


Figure 2. Evidence network



The size of each dot is proportional to the number of patients allocated and the width of line to the number of trials. Numbers do not add up to numbers in Table 1 because of a multi-arm trial by Propst1980.

Figure 3. Specific component of CBT for each number of sessions